Pre- and Short-term Posttreatment Vocal Functioning in Patients with Advanced Head and Neck Cancer Treated with Concomitant Chemoradiotherapy

Irene Jacobi1, Lisette van der Molen1, Maya van Rossum2, Frans Hilgers1,3,4

1The Netherlands Cancer Institute, Amsterdam, The Netherlands
2Leiden University Medical Center, Leiden, The Netherlands
3Institute of Phonetic Sciences, ACLC, University of Amsterdam, The Netherlands
4Academic Medical Center, University of Amsterdam, The Netherlands

i.jacobi@nki.nl

Abstract

Forty-seven patients with advanced larynx/hypopharynx, nasopharynx or oropharynx/oral cavity cancer were recorded before, and 10 weeks after concomitant chemoradiotherapy (CCRT), to investigate the effect of the tumor versus the effects of treatment. To evaluate voice functioning before and after treatment, voice quality and glottal behavior of sustained /a/ vowels were analyzed acoustically and compared with patient-based data on cigarette and alcohol usage. Acoustic measures of effort, nasality and regularity, such as periodicity or harmonics-to-noise ratio, differed significantly and progressed differently in dependence of the 3 distinct cancer/radiation sites. Baseline measures of voice stability correlated significantly with alcohol/smoking behavior.

Index Terms: pathological voice, cancer, chemoradiation

1. Introduction

In advanced head and neck cancers, treatment by chemoradiation has become a favored treatment modality. Concurrent chemo- and radiotherapy improves locoregional control and organ preservation, and increases survival [1]. Though improving local control and preserving the organ, the treatment can severely affect organ functioning and quality of life [2]. In head and neck cancers, next to survival, swallowing, nutrition and trismus have been the primarily reported outcome measures to indicate organ functioning and therewith the success of the treatment and the patient’s quality of life. However, also vocal functioning can be severely affected by the tumor site (larynx/pharynx) and its treatment by radiation (radiation fields also encompassing the larynx in case of oral/oropharyngeal and nasopharyngeal cancers). The combination with chemotherapy aggravates these effects. Common complaints in patients with laryngeal carcinoma are for example increased vocal effort, breathiness and hoarseness [3]. The tumor, often accompanied by significant edema, can distort voice quality by obstructing the airflow or impairing normal cord movement, whereas fibrosis, vocal fold atrophy, mucositis and swelling of tissues are effects of the treatment [4;5]. Next to this, treatment-induced changes in saliva production or its consistency also can affect voice quality.

Here, the voice of patients with advanced head and neck cancer was investigated by means of acoustic analysis prior to and after chemoradiotherapy (CRT), to assess effects of the tumor versus the effects of treatment.

2. Material and method

As part of a multidimensional assessment protocol of the randomized clinical trial “Prevention of trismus, swallowing and speech problems in patients treated with chemoradiation for advanced head and neck cancer” [6], the patient’s voice was recorded. Patient accrual for this so-called RADPLAT-protocol took place from September 2006 to April 2008.

2.1. Subjects

All included patients had an anatomical or functional inoperable primary tumor (advanced stage III or IV squamous cell carcinoma) of the oral cavity, oropharynx, hypopharynx, larynx, or nasopharynx. The original patient sample consisted of 55 patients, 44 males and 11 females. All patients were referred to the Netherlands Cancer Institute/ Antoni van Leeuwenhoek Hospital for their primary treatment and were eligible for concomitant chemoradiotherapy (CCRT) with curative intent. The patients had been introduced to swallowing exercises by a speech pathologist and were asked to train these before, during, and after treatment.

2.2. Recordings

For the voice analysis, the patients had been asked to produce sustained /a/-vowels at a comfortable loudness and pitch level. The recordings took place in a sound-treated room with a Sennheiser MD421 Dynamic Microphone and a portable 24bit digital wave recorder (Edirol Roland R-1) with a sampling frequency of 44.1 kHz. The mouth-to-microphone distance was kept constant at 30 cm.

Baseline and 10 weeks (median 10, range 9-11 weeks) posttreatment voice recordings were available for 47 of the 55 patients. The drop-out before the first follow-up was due to 3 deceases and 5 exclusions: two due to surgery, one had proven residue, one was unavailable, and one administrative miss. Table 1 on the following page shows the characteristics of the patients whose voice was available for acoustic analysis pretreatment and during the first follow-up.

2.3. Acoustic analysis

For acoustic measurements, one second of the most stable mid part of /a/ was selected, based on visual inspection of the spectrogram, and voicing in terms of nasality, noise, effort, stability and periodicity was assessed. The program PRAAT [7] was used for the following analyses:
2.3.1. Irregularity

Voice stability was defined by the ability to maintain a steady phonation, and was assessed in terms of jitter, shimmer, and the standard deviation of the fundamental frequency. For cycle-by-cycle perturbations of $f_0$, the average absolute difference between consecutive periods was related to the average $f_0$-period (jitter). Cycle-by-cycle perturbations of the amplitude were measured by relating the average absolute amplitude difference between consecutive periods to the average amplitude of the periods in the 1-second voice sample (shimmer). The standard deviation of the fundamental frequency in Hertz as a function of time was related to the speaker’s mean $f_0$ ($f_{0SD}$).

2.3.2. Spectral tilt

If more effort is required, this might be reflected in more high frequency energy, and a flatter spectral tilt. One spectral tilt measure was based on the difference in decibels between the mean spectral intensity in the frequency bands 500 to 4000 Hz and 4000 to 8000 Hz in the spectrum (STdB). Further band energy differences were assessed by relating the amplitude of the first harmonic to the amplitude of the third formant ($H1A3$), a spectral tilt measure correlating with glottal adduction and the bandwidth of the first formant [8]. The more abrupt vocal fold closure, the more energy in the higher frequencies. Since $H1A3$ correlates with the bandwidth of the first formant, it may as well be affected by nasalization.

2.3.3. Nasality

Nasalization can be indicated by an enhancement of the amplitude of the first harmonic, and a weakening and broadening of the first formant [8], as well as by decreased spectral energy in the area of the second and third formant. It was assessed by the difference between the amplitude of the first formant and the first harmonic ($H1A1$), and by the difference between the filter bands 2-3kHz and 3-4kHz (NAS).

Table 1: Patient characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>pre-/ 10 weeks posttreatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>47</td>
</tr>
<tr>
<td>Male</td>
<td>37</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
</tr>
<tr>
<td>Age</td>
<td>57 (32-78)</td>
</tr>
<tr>
<td>T category</td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>8</td>
</tr>
<tr>
<td>T2</td>
<td>15</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
</tr>
<tr>
<td>T4</td>
<td>6</td>
</tr>
<tr>
<td>N category</td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>4</td>
</tr>
<tr>
<td>N1</td>
<td>14</td>
</tr>
<tr>
<td>N2</td>
<td>24</td>
</tr>
<tr>
<td>N3</td>
<td>5</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>16</td>
</tr>
<tr>
<td>IV</td>
<td>31</td>
</tr>
<tr>
<td>Tumor site</td>
<td></td>
</tr>
<tr>
<td>Oral cavity, oropharynx</td>
<td>23</td>
</tr>
<tr>
<td>Nasopharynx</td>
<td>7</td>
</tr>
<tr>
<td>Laryngo-/hypopharynx</td>
<td>17</td>
</tr>
</tbody>
</table>

2.3.4. Noise components

The relative amount of high frequency noise was assessed by the band energy difference between 0 to 500 Hz and 4000 to 5000 Hz (BEDd)$. Noise components were also assessed by the harmonics-to-noise ratio (HNR), calculated by cross-correlation, and its standard deviation (HNRSD). As an alternative relative noise measure, the glottal-to-noise excitation ratio (GNE) was calculated. Other than HNR, this measure is not based on an estimation of $f_0$, and appears to be almost shimmer and jitter independent [9].

3. Results

3.1. Pretreatment

At baseline, prior to treatment, the patients’ daily units of alcohol correlated significantly with the instability of the fundamental frequency ($f_{0SD}$; $p<.001$, $r=.696$). The more alcohol units, the less stable $f_0$. Smoking in terms of cigarettes a day correlated weakly with $H1A3$, a measure for the speed of glottal fold closure ($H1A3$; $p=.005$, $r=.403$). The more cigarettes a day, the higher $A3$ compared to $H1$, and thus the more abrupt glottis closure. There was no correlation with age.

To assess possible effects by the tumor itself on phonation versus the effects of the treatment by chemoradiation, the patients were split according to tumor sites; nasopharyngeal cancer patients ($N=7$), laryngo/hypopharyngeal ($N=17$), and oral cavity/oropharynx cancer patients ($N=23$). When the patients’ pretreatment voice was evaluated by means of a Kruskal-Wallis test, the tumor group means differed significantly for nasality (NAS, $X^2(2)=7.565$, $p=0.023$), and marginally for the harmonics-to-noise-ratio (HNR, $X^2(2)=5.777$, $p=0.056$). There was also a significant effect for the daily units of alcohol ($X^2(2)=9.651$, $p=0.008$).

Mann-Whitney U-tests showed that the laryngo/hypopharyngeal cancer group consumed significantly more units of alcohol per day than the oral cavity/oropharyngeal cancer group ($U(39)=116.5$, $Z=-2.167$, $p=.030$) or the nasopharyngeal cancer group ($U(23)=16.500$, $Z=-2.740$, $p=.006$). Pretreatment, in most of the voice measures, the values of the patient group with nasopharyngeal cancer were in between the values of the other two groups as can be seen in figure 1 on the following page. There was no effect of age. Except for significantly more nasality in the nasopharyngeal cancer group ($H1A1$, $U(29)=35.000$, $Z=-2.231$, $p=.026$), there were no significant differences between the nasopharyngeal and the oral cavity/oropharyngeal cancer group. Mann-Whitney U-tests revealed that the nasopharyngeal cancer group showed also significantly more nasality than the patients with laryngo/hypopharyngeal cancer (NAS, $U(23)=7.565$, $Z=-2.117$, $p=.040$), and a significantly smaller standard deviation of the fundamental frequency ($f_{0SD}$, $U(23)=1.500$, $Z=-2.417$, $p=.013$). The mean value of the patients with laryngo/hypopharyngeal cancer indicated no nasality before treatment. However, the patient group with laryngo/hypopharyngeal cancer showed significantly worse harmonics-to-noise-ratios in the sustained /a/ (mean=14.171, $U(39)=110.500$, $Z=-2.326$, $p=.020$) than the patient group with oral cancer (mean=17.432, $U(39)=110.500$, $Z=-2.175$, $p=.030$), and a significantly smaller standard deviation (HNRSD). As an alternative relative noise measure, the glottal-to-noise excitation ratio (GNE) was calculated. Other than HNR, this measure is not based on an estimation of $f_0$, and appears to be almost shimmer and jitter independent [9].
Ten weeks posttreatment, independent of tumor site, paired t-tests showed that compared to pretreatment, values got significantly worse for the HNR (t-test p=.014, t=2.381), GNE (p=.015, t=2.736), and nasality (p=.043, t=2.200) got significantly worse after treatment. In the oral cavity/oropharyngeal group, the stability of the harmonics-to-noise ratio (HNRS) p=.026, t=2.379, BEDdB (p=.019, t=2.540), and measures of effort (spectral tilt p=.017, t=2.570, and h1a3, p=.003, t=3.291) worsened significantly. In the nasopharyngeal group, nasality/spectral tilt improved significantly (H1A3 p=.023, t=3.040).

4. Discussion

The results show that in patients with head and neck cancer, before treatment, voice measures reflected the patient’s drinking and smoking behavior. The higher standard deviations of f0 that were found for patients that reported

\[ X'(2) = 11.199, \ p = .004 \; \text{for jitter, } X'(2) = 10.400, \ p = .006 \; \text{for GNE, } X'(2) = 10.114, \ p = .006. \]
stronger alcohol consumption, could characterize a motor control problem [10]. H1A3 can indicate effort or less motor control as well. It correlated with the patients’ smoking behavior. Furthermore, pretreatment, the acoustic measures were significantly affected by the tumor (location). Velar incapability or changes in the (nas)o/phyngyngeal tract can lead to nasalization. The laryngo-hypopharyngeal group did not show any nasality prior to treatment. The oral/oropharyngeal group showed some nasality. However, not surprisingly, the highest nasality was found in the nasopharyngeal patient group prior to treatment, indicating that the nasopharyngeal tumor severely handicapped the closing-off of the nasal cavity during phonation. Whereas nasality decreased after treatment in the oral/oropharyngeal patient group and especially in the nasopharyngeal group, values increased in the laryngo/hypopharyngeal group. For the latter, nasalization therefore obviously is treatment-induced, whereas for the oral/oropharyngeal and nasopharyngeal cancer patients, nasalization was caused by the tumor and decreased strongly with tumor disappearance after treatment.

In view of treatment effects on voice, the outcome showed significant other differences depending on the tumor site. Several voice measures differed significantly between the laryngo/hypopharyngeal cancer patient group and the oral cavity/oropharyngeal cancer group. The least damage to voice quality by treatment indicate the values of the oral cavity/oropharyngeal patient group: A stronger spectral tilt and higher amounts of noise have been associated with breathiness [11,12], whereas jitter and shimmer have been associated with roughness and hoarseness. Jitter and shimmer may be related to changes in tissue properties, additive noise may depend on glottal leakage [11]. Only after treatment did the tumor groups differ significantly in jitter or shimmer. The laryngo/hypopharyngeal patient group showed the worst values. This could indicate that their tissues were damaged most severely when the three tumor groups are compared. It suggests furthermore that in this case the negative effects of chemo radiation on the laryngo-pharyngeal tissues outweigh the decrease of the tumor-induced changes.

Whereas the effect of treatment did not differ significantly between the tumor sites for e.g. f0SD, STdB, and HNR, there were significant tumor group differences in view of treatment effects for GNE, jitter and shimmer. For the oral/oropharyngeal cancer group, the GNE and jitter values suggest that periodicity and closure of the glottis remain unaffected by treatment. For the laryngeal/hypopharyngeal group, the values indicate aperiodicity and considerable glottal leakage after treatment. The nasopharyngeal group showed slightly worsened values as well.

An evaluation of the radiation doses, also in view of irradiated lymph nodes (in progress), will give further information on the effects of treatment and tumor location. Preliminary evaluations of mean doses indicate that differences from pre- to posttreatment in view of tumor location are due to differences in (local) radiation doses.

In general, the outcome shows that effects of cancer and short-term effects of its treatment on voice functioning differ in dependence of the tumor/radiation site; underlining that baseline data and a separation of laryngeal- and non-laryngeal cancers are mandatory in the evaluation of organ functioning after CCRT, as well as for an interpretation of tumor versus treatment effects. Acoustic voice analysis is an objective and helpful tool when analyzing organ functioning in head and neck cancer patients. The significant tumor and treatment effects on voicing call for more detailed investigations within the group of patients undergoing chemoradiotherapy. Future analysis will include the analysis of running speech. Furthermore, the outcomes will be compared to swallowing and trismus outcome that were assessed by patient questionnaires and by clinicians’ findings during the present trial.

5. References


